


Article

A Sentiment Analysis Approach for Exploring Customer Reviews of Online Food Delivery Services: A Greek Case

Nikolaos Fragkos¹, Anastasios Liapakis^{2,*}, Maria Ntaliani^{1,*}, Filotheos Ntalianis³ and Constantina Costopoulou¹

¹ Department of Agricultural Economics and Rural Development, School of Applied Economics and Social Sciences, Informatics Laboratory, Agricultural University of Athens, 75 Iera Odos St., 11855 Athens, Greece; tina@aua.gr (C.C.)

² Department of Archival, Library and Information Studies, School of Administrative, Economics & Social Sciences, University of West Attica, 28, Ag. Spyridonos St., 12243 Egaleo, Greece

³ Department of Business Administration, School of Economics, Business and International Studies, University of Piraeus, 80, M. Karaoli & A. Dimitriou St., 18534 Piraeus, Greece; filotheos@unipi.gr

* Correspondence: aliapakis@uniwa.gr (A.L.); ntaliani@aua.gr (M.N.)

Abstract: The unprecedented production and sharing of data, opinions, and comments among people on social media and the Internet in general has highlighted sentiment analysis (SA) as a key machine learning approach in scientific and market research. Sentiment analysis can extract sentiments and opinions from user-generated text, providing useful evidence for new product decision-making and effective customer relationship management. However, there are concerns about existing standard sentiment analysis tools regarding the generation of inaccurate sentiment classification results. The objective of this paper is to determine the efficiency of off-the-shelf sentiment analysis APIs in recognizing low-resource languages, such as Greek. Specifically, we examined whether sentiment analysis performed on 300 online ordering customer reviews using the Meaning Cloud web-based tool produced meaningful results with high accuracy. According to the results of this study, we found low agreement between the web-based and the actual raters in the food delivery services related data. However, the low accuracy of the results highlights the need for specialized sentiment analysis tools capable of recognizing only one low-resource language. Finally, the results highlight the necessity of developing specialized lexicons tailored not only to a specific language but also to a particular field, such as a specific type of restaurant or shop.

Keywords: sentiment analysis; customer reviews; online food delivery; food and beverage industry; Greece



Citation: Fragkos, N.; Liapakis, A.; Ntaliani, M.; Ntalianis, F.; Costopoulou, C. A Sentiment Analysis Approach for Exploring Customer Reviews of Online Food Delivery Services: A Greek Case. *Digital* **2024**, *4*, 698–709. <https://doi.org/10.3390/digital4030035>

Academic Editors: Phivos Mylonas and Yannis Manolopoulos

Received: 15 April 2024

Revised: 13 August 2024

Accepted: 16 August 2024

Published: 17 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sentiment analysis has gained substantial attention in recent years due to its potential for extracting meaningful insights from textual data. One application of sentiment analysis is within the domain of online food delivery. With the proliferation of online food platforms and the increasing use of social media for sharing experiences, understanding customer sentiments and feedback is crucial for the success and growth of these services [1].

Sentiment analysis in online food delivery has proven to be a useful tool for gaining an understanding of customers through their opinions, pinpointing areas that require development, and monitoring trends within this fast-paced sector. Through the analysis of customer reviews and feedback, sentiment analysis techniques contribute to enhancing customer satisfaction, optimizing service quality, and informing decision-making processes for online food delivery platforms [2].

There are two primary methodologies employed in sentiment analysis: machine-learning-based and lexicon-based approaches. Machine-learning-based sentiment analysis leverages algorithms to learn patterns and classify sentiment, while lexicon-based sentiment

analysis uses pre-defined sentiment lexicons. Both approaches have been widely applied in sentiment analysis tasks, including those in the food sector [3].

However, a significant challenge arises in the sentiment analysis of low-resource languages. Most of the available textual datasets for sentiment analysis are in English, while the analysis of low-resource languages poses many difficulties characterized by limited linguistic resources and complexities in grammar and vocabulary. The scarcity of available datasets in these languages hampers the automatic extraction of entities and sentiment classification. Due to this data deficiency, researchers working with low-resource languages either have to utilize the limited existing datasets or create their own [4,5].

The objective of this study is to evaluate the efficiency of off-the-shelf sentiment analysis APIs in recognizing and processing low-resources languages, like the Greek language. Given the limited availability of linguistic resources in the examined language, it is crucial to assess how effective the Meaning Cloud widely used API performs in terms of accuracy and reliability. Through this analysis, conclusions about consumer trends were drawn and the accuracy of a tool that first translates and then analyzes comments was measured. These results will help determine the effectiveness of low-resource language sentiment analysis tools.

Following this introduction, this paper unfolds into several key sections. The background section provides a theoretical underpinning of sentiment analysis techniques, which is followed by an overview of existing research on sentiment analysis in the food sector with a specific focus on low-resource languages. The methodology section outlines the approach followed, including data collection and the tool that was used in the analysis. In the results section, the findings of the analysis are presented along with key trends and assessment of the performance of the utilized sentiment analysis tool. Lastly, the conclusion summarizes the findings, the challenges and limitations of this work.

2. Background

2.1. Sentiment Analysis

Sentiment analysis is a subfield of natural language processing (NLP). It typically consists of three levels, which researchers have explored to gain a deeper understanding of this process and its applications in various domains. These are as follows:

- Document-level sentiment analysis focuses on the overall sentiment expressed in a document or a piece of text, such as a review, blog post, or social media post. This level of sentiment analysis provides a holistic view of the sentiment associated with the entire document. For example, Pang and Lee [6] conducted research on document-level sentiment analysis, employing machine learning techniques to classify movie reviews based on the overall sentiment expressed in the text.
- Sentence-level sentiment analysis focuses on analyzing the sentiment of individual sentences within a document. It aims to determine the sentiment polarity (positive, negative, or neutral) of each sentence. This level of sentiment analysis allows for a more fine-grained understanding of sentiment within a document. For instance, Socher and colleagues [7] proposed a recursive neural network model for sentence-level sentiment analysis, achieving state-of-the-art performance on sentiment classification tasks.
- Aspect-level sentiment analysis focuses on extracting sentiment associated with specific aspects or entities mentioned in the text. It aims to identify the sentiment polarity for different aspects mentioned within a document, allowing for a more detailed analysis. For example, Wang and colleagues [8] proposed a novel neural network-based approach for aspect-level sentiment analysis, which was able to effectively capture sentiment information related to specific aspects in user reviews.

These three levels of sentiment analysis provide researchers and practitioners with different perspectives on sentiment understanding, enabling them to gain insights at various granularities. By employing techniques at these levels, sentiment analysis can be effectively applied in fields, such as customer feedback analysis, social media monitoring, and em-

ployee and market research, among others. There are three commonly used approaches in sentiment analysis [9–13]:

- A machine-learning-based approach involves training models on labeled data to automatically classify sentiment in text. This approach uses algorithms, such as support vector machines (SVM), random forests, and neural networks to learn patterns and features indicative of sentiment. In the abovementioned work of Pang and Lee [6], movie reviews were classified as positive or negative by employing a machine learning approach, namely, an SVM classifier. This approach has also been applied in the food sector for food recognition and classification; more specifically, deep learning is used for food quality detection and food safety in food supply chain [14].
- A lexicon-based approach relies on predefined sentiment lexicons or dictionaries to determine the sentiment polarity of text. It involves assigning sentiment scores to individual words or phrases based on their presence in the lexicon, which contains a list of words annotated with their associated sentiment polarities (e.g., positive, negative, or neutral). This approach estimates the overall sentiment expressed in a given text by using the semantic orientation of words. One widely used lexicon-based approach is the Valence Aware Dictionary and Sentiment Reasoner (VADER) lexicon. VADER utilizes a comprehensive sentiment lexicon that incorporates both polarity (positive/negative) and intensity (strength) of sentiment words. It also accounts for the influence of contextual valence shifters (e.g., “but”, “however”) and punctuation in sentiment analysis [15].

However, a lexicon-based approach may face challenges when encountering words or phrases that are not present in the lexicon or when dealing with sarcasm, irony, or other forms of contextual sentiment expression [16]. Despite these limitations, this approach has been widely applied in sentiment analysis tasks across various domains, including social media, product reviews, and customer feedback analysis. In the food sector, lexicon-based sentiment analysis has been applied to analyze customer sentiments toward food trends. For example, Twitter posts were analyzed in order to detect differences between geographical region regarding new food trends [17].

- A hybrid approach comprises the amalgamation of the abovementioned approaches. Machine-learning-based approaches offer flexibility and adaptability, while lexicon-based approaches provide simplicity and interpretability. In an effort to achieve better results, researchers are exploring the potential of the combination of various approaches and tools. They continue to refine sentiment lexicons and develop hybrid approaches that combine machine-learning-based and lexicon-based approaches with other techniques to improve sentiment analysis accuracy, applicability, and robustness in different contexts. Such is the work of Appel and colleagues (2018), proposing a hybrid approach that uses NLP essential techniques, a sentiment lexicon enhanced with ‘SentiWordNet’, and fuzzy sets to determine the semantic orientation polarity and its intensity for sentences [18].

2.2. Sentiment Analysis in the Food Sector

This section provides an overview of key studies, including research papers on the implementation of sentiment analysis in the food sector with a specific focus on low-resource languages. The studies included were confined to those published in English between January 2011 and December 2023. For our research purposes, the following databases were used: Scopus, Willey Online Library, and Web of Science.

The studies included in this literature review explored sentiment analysis in the context of online food delivery to gain insights into customer experiences and satisfaction.

Khan and colleagues [19] conducted a study on sentiment analysis of online food delivery review and identified that price and hygiene affect the sentiment of the customer towards online food delivery platforms. Similarly, Liu and colleagues [20] underlined the importance of easy payment, customization and fast delivery. Their analysis of customer preferences for online shopping was based on optimized feature extraction using the

Principal Component Analysis with a Social Spider Optimization (PCA-SSO) algorithm. The gathered data were in the English language and the aim of the study was to improve food service quality in online shopping and offer insights regarding customer satisfaction. Teichert and colleagues [21] conducted a multi-dimensional analysis by gathering feedback about food delivery services and analyzing it along two axes. The axes were the actual product, which includes product issues and brand satisfaction; and the augmented product, including payment process and service handling. Employing web scraping, text mining, and multivariate statistics analysis, their aim was to understand the consumer experience on dimensions crucial for business success.

Moreover, sentiment analysis has been utilized to help online food delivery companies gain competitive advantage through their customer-generated content of social media. The findings emphasized the polarity of the content and recommendations for business to change this polarity [22]. Vatambeti and colleagues [23] collected and analyzed consumer posts from Twitter about online food delivery services like Swiggy, Zomato, and UberEATS. The researchers utilized a combination of Convolutional Neural Network (CNN) and Bi-directional Long Short-Term Memory (Bi-LSTM) models. The study aimed not only to assess the performance of the models but also to gain insights into consumer sentiment towards these three platforms.

Adak and colleagues [3] conducted sentiment analysis on comments from food delivery services (FDS) like UberEATS and Deliveroo with the aid of deep learning models. Despite achieving high-performance metrics, these models lack computing transparency. To address this issue, they employed explainable Artificial Intelligence (AI) techniques such as Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP), aiming to increase the interpretability of decisions made by these models. The output of their research was the accurate classification of comments in order to address issues and improve customer satisfaction in the FDS domain.

In a different linguistic context, Nguyen and colleagues [1] delved into the analysis of sentiment in customer feedback in online food ordering services, focusing on reviews in the Vietnamese language. They extracted 236,867 reviews, and employed four lexicon-based models and a support vector machine from scratch, which achieved the higher evaluation metrics. The goal of the study was twofold: to generate an accurate model, and to provide insights about the top stores and sentiment trends over time. Altaf and colleagues [24] explored cross-domain sentiment analysis in Urdu, an underexplored research area for low-resource languages. Their baseline proposed method involved the use of n-grams and word embedding along with machine learning and deep learning classifiers. The study aimed to evaluate the performance of a domain-specific classifier in a different domain achieving a relatively high F1 score on the cross-domain application. Similarly, Zufiker and colleagues [25] proposed a sentiment analysis approach for Bangla texts using deep learning algorithms, emphasizing the limitations of analyzing low-resource languages like Bangla. The authors analyzed texts from the e-commerce platform Daraz by leveraging deep learning techniques such as variant Convolutional Neural Network that outperforms the conventional machine learning techniques.

Shifting the focus to the impact of external events, Jang and colleagues [26] analyzed social media posts which related to food delivery before and after the COVID-19 outbreak. Utilizing Ucinet6 for detecting meaningful relationships among keywords and CONCOR analysis for sentiment network analysis, the researchers observed a small decrease in positive comments and slight increase in negative ones.

Concluding the series of research studies, Kumar and colleagues [27] collected 27,337 online customer reviews from a grocery shopping app to identify factors of consumer satisfaction. They applied Latent Dirichlet Analysis followed by correspondence analysis with the goal of contributing to customer satisfaction management.

This literature review highlights the significance of sentiment analysis in the food sector for understanding customer sentiments, enhancing service quality, monitoring trends, and assessing brand perception. Through the analysis of customer reviews, social media

data, and other textual sources, sentiment analysis provides valuable insights that can aid decision-making, marketing strategies, and overall customer satisfaction in the food industry. Combinations of various methods are used. However, the focus on low-resource languages is very low, as only two out of the twelve studies included in this review focused on this domain, without an exact application on the online food ordering sector. Despite the widespread use of advanced specialized tools for sentiment analysis, none of the studies evaluated the performance of specific tools that now require knowledge of coding, like Meaning Cloud.

3. Materials and Methods

This section presents the tool used, namely Meaning Cloud, as well as the steps followed for conducting sentiment analysis on comments posted by Greek consumers on an online food delivery platform.

3.1. The Meaning Cloud Tool

Meaning Cloud is a tool that provides an Application Programming Interface (API) where the user may choose the language of the output as well as the language of the inserted content. When the user selects either 'raw' or 'formatted' results, the analysis begins once some parameters have been set (Meaning Cloud Documentation. Available online: <https://learn.meaningcloud.com/developer/sentiment-analysis/2.1/doc/request#model> (accessed on 27 June 2023)). The algorithm behind Meaning Cloud's Sentiment Analysis API employs a combination of natural language-processing (NLP) techniques, machine learning models, and linguistic rules to evaluate the sentiment of a given text. Initially, the input text undergoes preprocessing to clean and normalize it. The procedure involves steps such as tokenization, where the text is broken down into individual words or phrases, removing stop words, which are common words that do not carry significant meaning, like articles, quantitative determinations, punctuation marks, etc., and stemming or lemmatization, which reduces words to their base or root form. Following preprocessing, the algorithm performs syntactic and semantic analysis. Syntactic analysis helps to determine the grammatical structure of the sentences, identifying parts of speech and the relationships between words. Semantic analysis aims to grasp the meaning of the words and phrases in the context of the sentence, including recognizing named entities such as the names of people, organizations, and locations, and understanding the context in which words are used. Using predefined linguistic rules and machine learning models, the algorithm then detects the polarity of the text. This involves assessing whether the sentiment expressed is positive, negative, neutral, or mixed. The algorithm can also provide detailed insights into different aspects or topics within the text, giving a more nuanced understanding of the sentiments related to specific entities or aspects mentioned in the text. The Greek language is supported by the Meaning Cloud's Sentiment Analysis API. However, it has to be mentioned that because of the limited resources available, most human languages, aside from English, are regarded as low-resource languages, which makes it difficult to automate information extraction tasks. Because of this issue, there is a noticeable discrepancy in the accuracy of existing natural language-processing tools for rich languages and low-resource languages. The parameters of the analysis are explained thoroughly below:

- "Verbose", more information is provided about the analysis and different polarities of the entities are detected.
- "Model" is the default sentiment model which is used for the analysis but there is also an option for the user to upload his own model.
- "Relaxed Typography" indicates how reliable the text to analyze is (as far as spelling, typography, etc., are concerned), and influences how strict the engine will be when it comes to taking these factors into account in the analysis.
- "Expand Global Polarity" allows us to choose between two different algorithms for the polarity detection of entities and concepts. Enabling the parameter gives less weight

to the syntactic relationships; so, it is recommended for short texts with unreliable typography.

- “Guess unknown words” adds a stage to the sentiment analysis in which the engine tries to find a suitable analysis to the unknown words resulting from the initial analysis assignment. It is especially useful to decrease the impact typos have in text analyses.
- “Disambiguation level” contains the semantical and morphosyntactic disambiguation in order to determine the meaning of a word or its specific usage in a particular sentence.

The score tag in the coding format displays the comment’s overall polarity, which, in this instance, is strongly positive. The polarities detected by the tool are shown below:

- No Polarity—NONE.
- Strong Negative—N+.
- Negative—N.
- Neutral—NEU.
- Positive—P.
- Strong Positive—P+.

First, the text is inserted into the Text Box for each comment analysis. The analysis’s findings are then displayed along with the aspects that were found and registered in an Excel file.

More precisely, the final score is the sum of the individual scores of the aspects. For example, any aspect with a positive polarity increases the score by one (1), while a neutral aspect has no effect on the score and a negative aspect lowers the score by one (1). Specifically, each of the analyzed comments with a score higher than zero (0) is classified as positive, each comment with a score lower than zero (0) is classified as negative and each comment with a score equal to zero (0) is classified as neutral.

3.2. Sentiment Analysis for Online Food Delivery

The steps followed for undertaking sentiment analysis on text reviews and comments that were uploaded on the “e-food” platform (E-food. Available online: <https://www.e-food.gr/> (accessed on 15 May 2023))—the most dominant online food delivery platform by offering access to 20,000 stores in 100 cities in Greece—are as follows:

- Firstly, the comments were mined using the “Data Scraper” tool, a Google Chrome extension (Data Miner. Available online: <https://dataminer.io/> (accessed on 10 March 2023)).
- The mined comments were inserted into the Meaning Cloud (Meaning Cloud. Available online: <https://www.meaningcloud.com/> (accessed on 27 June 2023)) tool to determine their overall polarity. This tool was chosen since it supports the Greek language by translating the texts that are inserted into it rather than by using a specialized lexicon. The results of the tool reviews were stored in an Excel file.
- To examine the efficiency of the Meaning Cloud SA API in Greek Language, the same comments were manually reviewed by two annotators (Table 1) that assigned a positive or negative sentiment polarity based on their personal judgment. The results of the annotators’ reviews were also stored in the Excel file.

Table 1. The annotators’ profiles.

	Annotator 1	Annotator 2
Level of Education	High school graduate, undergraduate student in B (Eng) Agricultural Economics and Rural Development	Post Doc (Informatics), PhD (Informatics), MBA, MSc, B (Eng)
Work	Student	University Adjunct Assistant Professor
Languages	Mother tongue Greek, proficient-level English	Mother tongue Greek, proficient-level English

The principal purpose is to gauge the tool's accuracy in analyzing the Greek language, which is not performed directly but rather by translating the text first. By comparing the tool's results to those of the experts', we must calculate four metrics, 'Accuracy', 'Precision', 'Recall' and 'F-score', based on the confusion matrix (Table 2) [28]. For the calculation of the metrics, the terms 'True Positive' and 'True Negative' were used. A comment is labeled as 'True Positive' when it is actually positive and the tool predicts its label as positive as well, and the same procedure is applied for the labeling of the negative comments. If the prediction made by the tool matches the actual labeling of the comment, it is correctly labeled and termed 'True'. However, if the prediction and the actual labeling differ, the comment is inaccurately labeled and termed 'False'. To facilitate comparison between the aspects identified by the experts and those identified by the tool, both sets were recorded in an Excel file. Based on the aspects that are used more frequently, certain consumer behavior inferences can be drawn.

Table 2. Confusion matrix.

	Predicted Positive	Predicted Negative	Total
Actual Positive	True Positive (tp)	False Negative (fn)	Total Positive
Actual Negative	False Positive (fp)	True Negative (tn)	Total Negative

Using the confusion matrix (Table 2), the abovementioned metrics are calculated as shown in Equations (1)–(4) [29].

$$\text{accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}} \quad (1)$$

$$\text{precision (p)} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad \text{precision (n)} = \frac{\text{tn}}{\text{tn} + \text{fn}} \quad (2)$$

$$\text{recall (p)} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad \text{recall (n)} = \frac{\text{tn}}{\text{tn} + \text{fp}} \quad (3)$$

$$\text{F - score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

(iv) Then, an entity analysis was undertaken by the experts and the tool. The following entities were proposed to undergo sentiment analysis:

- "Price": regards the pricing of the order.
- "Speed": refers to the delivery time of the order.
- "Quality": concerns the overall quality of the order.
- "Behavior": refers to the delivery personnel's behavior.
- "Hygiene": regards the restaurant's hygiene.
- "Overall impression": concerns the restaurant's overall image.
- "Portion size": refers to the portion size of the order.
- "Service": regards the customer service received by the restaurant.

4. Results

Three hundred (300) comments were collected. These comments were collected by three (3) different types of food-related businesses, a hundred (100) comments from each one. A fast-food restaurant, an Italian restaurant, and a coffee roaster shop comprised the three categories. These three categories were chosen because they are completely different in how they observe how the tool reacts to different vocabularies concerning each category. The total count of comments analyzed was 293 since some of them were eliminated from the analysis to avoid duplications and comments written in any language other than Greek. The number of analyzed comments from the fast-food restaurant, the Italian restaurant and the coffee roaster shop was 98, 98 and 97, respectively.

Overall, the analysis shows high performance in the classifications of the dataset, with an average accuracy of 90.67% (Table 3). It should be underlined that 34% (100 comments) of the dataset was not classified from the model, 25% (75) was not evaluated due to sarcasm and lacking syntax, and 8% (25) was classified as neutral. Essentially, the model classified only 65% of the comments, namely 193 out of 293. The observed percentage can be attributed to the nature of the Greek language. The Greek language is characterized by a complex grammar, vocabulary and syntax. Also, the comments were not capitalized but they were analyzed raw after being extracted from the platforms, which meant some words may not have been identified by the tool due to their different accentuation. These words may have the same meaning but a different orthographic representation due to the difference in their accent marks [28]. Additionally, the tool itself first translates the text and then conducts the sentiment analysis classification. This two-step procedure adds a layer of complexity and a margin of error. During this translation, some words may not be translated correctly, resulting either in wrong classification or no classification at all if the translated word is not recognized by the tool. Also, the tool performs analysis at a sentence level not at an aspect level, which is not preferable because customers' reviews in the food and beverage domain evaluate various aspects. The consumer may address different aspects for an order, but all these aspects will be aggregated together, and they will not be examined separately. This fact compromises the granularity, and the specificity of the analysis are compromised. Moreover, there is a positive trend towards online food ordering as the true positive comments were one hundred and thirty-five (135), almost three (3) times the negative ones, as shown in the confusion matrix (Table 4).

Table 3. Overall performance of the system in the dataset.

	Total Positive	Total Negative
Precision	91.12%	88.88%
Recall	96.42%	75.47%
F-Score	93.70%	81.62%
Accuracy	90.67%	

Table 4. Overall confusion matrix.

	Predicted Positive	Predicted Negative	Total	Actual
Actual Positive	135	5	140	179
Actual Negative	13	40	53	94

Table 5 statistically examines whether there is agreement between the experts and the tool used. It specifically presents the interrater agreement results between the trained experts and the Meaning Cloud text analytics platform, as well as the Intraclass Correlation (ICC) for the combined data and each individual company. As regards Cohen's unweighted kappa, the values range between 0.25 and 0.28 for all Greek companies and the combined data, which shows a fair agreement between the expert and the Meaning Cloud tool [29]. Similarly, Fleiss' kappa shows a fair agreement for all data analyzed. Finally, Krippendorff's alpha, a more conservative test, shows a tentative agreement [29] only for the fast-food company (0.67) but not for the other companies and the total dataset. As regards the ICCs, all are above the acceptable benchmark values [30,31]. Overall, our data show a marginal agreement between the ratings of the experts and those of the tool.

Table 5. Interrater agreement indicators and intraclass correlations for the three Greek companies.

	Cohen's κ_w	95% CI	Fleiss' κ	95% CI	Krippendorff's α	95% CI	ICC
Fast Food	0.26	0.12-0.39	0.25	0.11-0.38	0.67	0.11-37	0.72
Italian Restaurant	0.25	0.11-0.39	0.22	0.09-0.36	0.62	0.06-0.37	0.73
Coffee shop	0.28	0.12-0.45	0.28	0.13-0.43	0.63	0.44-0.45	0.74
Total	0.27	0.18-0.36	0.26	0.18-0.34	0.64	0.55-0.71	0.74

Note: N = 293, κ_w = kappa unweighted, κ = kappa, α = alpha, Total = combined data for all three companies, ICC = intraclass correlation.

Furthermore, in order to draw conclusions about the online food ordering from the consumer's perspective, an entity analysis was also carried out. As shown in Table 6, which presents the results from the expert's entity analysis, 8.2% of the comments regarded price, 44.7% addressed delivery speed, and 51.8% pertained to the quality of orders. In addition, 11.6% concerned the delivery personnel's behavior, 7.5% regarded hygiene, and 15.3% discussed the restaurant's overall impression. Finally, 6.8% concerned portion size, while 23.5% were focused on customer service. Overall, it is noticeable that consumers prioritize overall quality when evaluating their online food-ordering experience, followed by the speed of the order and the overall impression that the restaurant leaves. The speed of the order has a pivotal role in shaping the overall experience of the consumer, as any delays can lead to dissatisfaction and erosion of consumer trust in regard to the specific restaurant. The overall impression of a restaurant could be also considered as existing bias as the consumer already knows the restaurant and has set a specific standard.

Table 6. Results from the experts' entity analysis.

	Fast-Food	Italian	Coffee Roaster Shop	Total	Percentage
Price	6	12	6	24	8.2%
Speed	38	44	49	131	44.7%
Quality	56	46	50	152	51.8%
Behavior	6	15	13	34	11.6%
Hygiene	11	9	2	22	7.5%
Overall Impression	18	20	7	45	15.3%
Portion Size	15	5	0	20	6.8%
Service	21	30	18	69	23.5%

Table 7 presents the entity analysis results of the tool. The tool failed to detect comments regarding the entities of price, hygiene, and portion size. Concerning the rest of the entities, the tool identified comments as follows: 0.7% about speed, 31.7% about quality, 4% about the delivery personnel's behavior, 13.6% about the restaurant's overall impression, and 10.9% about the service. Lastly, in 39.9% of all comments, the tool failed to identify any entities, while 8.8% contained isolated labels covering various food categories, amalgamated into the 'other' entity. The efficiency of the tool is directly linked to the words already included in the default dictionary used for the analysis. Specifically, the words with 0% coverage were not included at all. Additionally, the significant proportion of comments that were not identified can be attributed to the fact that semantic value may be lost during translation. This highlights the need for further enhancement of the dictionary by encompassing a broader range of relevant terms and categories.

Table 7. Results from the entity analysis of the tool.

	Fast-Food	Italian	Coffee Roaster Shop	Total	Percentage
Price	0	0	0	0	0%
Speed	1	0	1	2	0.7%
Quality	30	38	25	93	31.7%
Behavior	2	7	3	12	4%
Hygiene	0	0	0	0	0%
Overall Impression	17	17	6	40	13.6%
Portion Size	0	0	0	0	0%
Service	9	14	9	32	10.9%
None	31	30	56	117	39.9%
Other	15	6	5	26	8.8%

The analysis highlights gaps in the detection capabilities of the tool since it reveals an inconsistency between the experts' observations and the findings of the tool. The expert identified quality, speed, and customer service as the most pivotal entities, and all the entities were included in the observations. The tool primarily detected comments on overall impression, quality, and customer service, while price and hygiene were missing from the observations.

5. Conclusions

Sentiment analysis plays a crucial role in understanding the consumers' thoughts towards products, services, and purchasing experiences. By analyzing sentiment, businesses can understand customers' emotional requirements and make decisions that nurture deeper connections with their customer base.

This study contributes to the growing body of literature on sentiment analysis in online food delivery services, providing evidence from Greece that underscores the importance of understanding customer sentiment for the success and growth of these platforms. It tried to investigate the effectiveness of off-the-shelf sentiment analysis APIs in providing meaningful and accurate results for identifying sentiments in Greek.

According to the results, the analysis achieved a high accuracy level. The tool detected correctly 76% of the positive comments and 42% of the negative comments. There are three times more positive comments than negative ones. It must be noted that although the classification had a high accuracy, only 66% of the total comments were classified. Therefore, if the unclassified comments were included in the evaluation metrics, the percentages would probably decrease. However, larger datasets in forthcoming research should be used to increase the generalizability and robustness of the conclusions drawn.

Also, the findings of this research underscore that, when ordering food online, customers make comments mainly on the quality of the delivered meal, the speed of delivery, and the restaurant's customer service. By leveraging sentiment analysis techniques, we identified key points of customer interest. This insight can assist companies operating Greek food delivery platforms and the collaborating food catering businesses in improving customer satisfaction and optimizing service delivery.

It must be noted that, apart from the term 'store', which was incorporated into the generalized-default model utilized in the analysis, the percentages of discovered entities by the tool were quite low. The low percentages primarily stemmed from the research constraint requiring comments to be translated before analysis. This process often distorts the original meaning of the comments.

Moreover, this study has revealed challenges associated with the need for developing specialized tools tailored to the linguistic nuances of specific languages. As shown, the model used lacks specialization in a specific domain to incorporate relevant vocabulary;

instead, it relies on a limited set of terms in a generalized manner. This leads to the necessity of developing a lexicon dedicated not only to a specific language, but also to a particular field, in this case, a particular type of restaurant or shop. By developing the lexicon, the percentages will undoubtedly increase, yet achieving 100% accuracy is improbable due to customers employing incorrect or unstructured syntax. Such variations alter the meaning of comments, affecting the findings of the analysis. Moving forward, continued research and innovation in sentiment analysis tools and techniques will be essential for unlocking its full potential in diverse linguistic contexts and industry domains. Moreover, future work should include the deployment of other off-the-shelf sentiment analysis APIs and the comparison of their results with the findings of this work.

Author Contributions: Conceptualization, C.C., A.L., M.N., F.N. and N.F.; methodology, C.C., M.N., F.N. and A.L.; validation, N.F., A.L. and F.N.; formal analysis, N.F. and F.N.; investigation, N.F. and F.N.; resources, N.F.; data curation, N.F. and F.N.; writing—original draft preparation, N.F., M.N., N.F., C.C. and A.L.; writing—review and editing, N.F., M.N., N.F., C.C. and A.L.; visualization, N.F. and F.N.; supervision, M.N., N.F., C.C. and A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were created and analyzed in this study. These data are openly available here (in Greek): <https://informatics.aua.gr/research/datasets/> (accessed on 27 June 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nguyen, B.; Nguyen, V.-H.; Ho, T. Sentiment analysis of customer feedback in online food ordering services. *Bus. Syst. Res. J.* **2021**, *12*, 46–59. [CrossRef]
2. Shaeali, N.S.; Mohamed, A.; Mutalib, S. Customer reviews analytics on food delivery services in social media: A review. *IAES Int. J. Artif. Intell. (IJ-AI)* **2020**, *9*, 691. [CrossRef]
3. Adak, A.; Pradhan, B.; Shukla, N. Sentiment analysis of customer reviews of food delivery services using deep learning and explainable artificial intelligence: Systematic review. *Foods* **2022**, *11*, 1500. [CrossRef] [PubMed]
4. Magueresse, A.; Carles, V.; Heetderks, E. Low-resource languages: A review of past work and future challenges. *arXiv* **2020**, arXiv:2006.07264.
5. Aivatoglou, G.; Fytali, A.; Arampatzis, G.; Zaikis, D.; Stylianou, N.; Vlahavas, I. End-to-end aspect extraction and aspect-based sentiment analysis framework for low-resource languages. *Intell. Syst. Conf.* **2024**, *824*, 841–858. [CrossRef]
6. Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends® Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
7. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. *EMNLP* **2013**, *1631*, 1631–1642.
8. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for Aspect-level Sentiment Classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615. [CrossRef]
9. Birjali, M.; Kasri, M.; Beni-Hssane, A. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowl. Based Syst.* **2021**, *226*, 107134. [CrossRef]
10. Madhoushi, Z.; Hamdan, A.R.; Zainudin, S. Sentiment analysis techniques in recent works. In Proceedings of the 2015 Science and Information Conference (SAI), London, UK, 28–30 July 2015; pp. 288–291. [CrossRef]
11. Thakkar, H.; Patel, D. Approaches for Sentiment Analysis on Twitter: A State-of-Art study. *arXiv* **2015**, arXiv:1512.01043.
12. Nasim, Z.; Rajput, Q.; Haider, S. Sentiment analysis of student feedback using machine learning and lexicon based approaches. In Proceedings of the 2017 International Conference On Research And Innovation In Information Systems (ICRIIS), Langkawi, Malaysia, 16–17 July 2017; pp. 1–6. [CrossRef]
13. Sadia, A.; Khan, F.K.; Bashir, F. An Overview of Lexicon-Based Approach For Sentiment Analysis. 2018. Available online: <https://api.semanticscholar.org/CorpusID:201105314> (accessed on 24 January 2024).
14. Zhou, L.; Zhang, C.; Liu, F.; Qiu, Z.; He, Y. *Application of Deep Learning in Food: A Review*; Blackwell Publishing Inc.: Oxford, UK, 2019. [CrossRef]
15. Rintyarna, B.S. Mapping acceptance of Indonesian organic food consumption under COVID-19 pandemic using sentiment analysis of twitter dataset. *J. Theor. Appl. Inf. Technol.* **2021**, *99*, 1009–1019. Available online: <https://www.jatit.org/> (accessed on 27 January 2024).

16. Polignano, M.; Basile, V.; Basile, P.; Gabrieli, G.; Vassallo, M.; Bosco, C. A hybrid lexicon-based and neural approach for explainable polarity detection. *Inf. Process Manag.* **2022**, *59*, 103058. [CrossRef]
17. Pindado, E.; Barrera, R. Using Twitter to explore consumers' sentiments and their social representations towards new food trends. *Br. Food J.* **2021**, *123*, 1060–1082. [CrossRef]
18. Appel, O.; Chiclana, F.; Carter, J.; Fujita, H. A hybrid approach to sentiment analysis with benchmarking Results. In Proceedings of the Trends in Applied Knowledge-Based Systems and Data Science: 29th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, Morioka, Japan, 2–4 August 2016; pp. 242–254. [CrossRef]
19. Khan, F.M.; Khan, S.A.; Shamim, K.; Gupta, Y.; Sherwani, S.I. Analysing customers' reviews and ratings for online food deliveries: A text mining approach. *Int. J. Consum. Stud.* **2023**, *47*, 953–976. [CrossRef]
20. Liu, W.; Alqhatani, A.; Asiri, F.; Salwana, E. Customer preference analysis towards online shopping decisions based on optimized feature extraction. *Expert Syst.* **2023**, 1–15. [CrossRef]
21. Teichert, T.; Rezaei, S.; Correa, J.C. Customers' experiences of fast food delivery services: Uncovering the semantic core benefits, actual and augmented product by text mining. *Br. Food J.* **2020**, *122*, 3513–3528. [CrossRef]
22. Trivedi, S.K.; Singh, A. Twitter Sentiment Analysis of App Based Online Food Delivery Companies. *Glob. Knowl. Mem. Commun.* **2021**, *70*, 891–910. Available online: <https://api.semanticscholar.org/CorpusID:233967660> (accessed on 28 January 2024). [CrossRef]
23. Vatambeti, R.; Mantena, S.V.; Kiran, K.V.D.; Manohar, M.; Manjunath, C. Twitter sentiment analysis on online food services based on elephant herd optimization with hybrid deep learning technique. *Clust. Comput.* **2024**, *27*, 655–671. [CrossRef]
24. Altaf, A.; Anwar, M.W.; Jamal, M.H.; Hassan, S.; Bajwa, U.I.; Choi, G.S.; Ashrafet, I. Deep learning based cross domain sentiment classification for urdu language. *IEEE Access* **2022**, *10*, 102135–102147. [CrossRef]
25. Zulfiker, S.; Chowdhury, A.; Roy, D.; Datta, S.; Momen, S. Bangla E-Commerce Sentiment Analysis Using Machine Learning Approach. In Proceedings of the 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (sti), IEEE, London, UK, 28–30 July 2015; pp. 1–5. [CrossRef]
26. Jang, J.; Lee, E.; Jung, H. Analysis of food delivery using big data: Comparative study before and after COVID-19. *Foods* **2022**, *11*, 3029. [CrossRef] [PubMed]
27. Kumar, A.; Chakraborty, S.; Bala, P.K. Text mining approach to explore determinants of grocery mobile app satisfaction using online customer reviews. *J. Retail. Consum. Serv.* **2023**, *73*, 103363. [CrossRef]
28. Liapakis, A.; Tsiligiridis, T.; Yialouris, C. A sentiment lexicon-based analysis for food and beverage industry reviews. *Greek Lang. Paradigm. Int. J. Nat. Lang. Comput.* **2020**, *9*, 21–42. [CrossRef]
29. Landis, J.R.; Koch, G.G. The measurement of observer agreement for categorical data. *Biometrics* **1977**, *33*, 159. [CrossRef]
30. Koo, T.K.; Li, M.Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* **2016**, *15*, 155–163. [CrossRef] [PubMed]
31. Cicchetti, D.V. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess* **1994**, *6*, 284–290. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.